Pisco Med Publishing

Original Research Article

A Wealth of Data

Lei Xu , Zekun Li

School of Institute of Information Technology of Guet (Guilin 541000)

Abstract: According to the demand of Amazon online marketplace, we screened the data for completeness and redundancy. The star-rating and review-number were selected as independent variables, and helpful-votes was the dependent variable. Then we analyze the sentiment index of all comments through NLP, and use SPSS to analyze the time series after standardized processing.

Keywords: Correlation analysis; Multiple regression linear model; Time series analysis

The purpose of writing this article is to analyze the data set of three new products launched and sold by Sunshine Company online, and use mathematical evidence to observe whether these data help evaluate Sunshine Company's star-rating, review-number and helpful-votes Whether it will succeed in three new online marketplace products.

1 Analysis of Specific Issues

For the problem, It requires the use of product star-rating, review-number and helpful-votes to analyze quantitative or qualitative models, relationships, metrics or evaluations, and various parameters. After searching for literature and analyzing and processing data, we decided to use three variables: Establish an evaluation standard system with "input" and "output" by evaluating the star-rating, the number of comments and the number of comments in favor of the review, and using correlation analysis and multiple regression analysis to find the correlation coefficient between variables.

2 Model establishment and solution

The main purpose of Model is to study the correlation between the respective variables and the dependent variables, and use the correlation analysis to make a preliminary judgment. Correlation analysis is a commonly used statistical method to study the correlation between random variables. Correlation coefficient r is used to indicate the degree of correlation between two variables, and sample data is used to calculate the value of r, which ranges from -1 to 1. r> 0 means there is a positive correlation between the two variables, otherwise it is a negative correlation. r = 0 means there is no correlation between variables. | r |> 0.8 indicates that there is a strong correlation between variables, and | r | <0.3 indicates that the correlation between variables is very weak and can be considered irrelevant.

Through the above model, we used SPSS to correlate the standardized data with the star- rating and review-number and helpful-votes, and obtained correlation coefficients among three variables in the three markets. Table 1 and table 2 and table 3shows.

Table 1 Hair dryer correlation

		star-rating	review-number	helpful-votes
star-rating	Pearson Correlation	1	-.103**	-.029**
	Sig.(2-tailed)		.000	.004
	N	9766	9766	9766
review-number	Pearson Correlation	-.103**	1	.260**
	Sig.(2-tailed)	.000		.000
	N	9766	9766	9766
helpful-votes	Pearson Correlation	-.029**	.260**	1
	Sig.(2-tailed)	.004	.000	
	N	9766	9766	9766

Table 2 Microwave oven correlation

		star-rating	helpful-votes	review-number
star-rating	Pearson Correlation	1	.011	-.169
	Sig.(2-tailed)		.650	.000
	N	1615	1615	1615
helpful-votes	Pearson Correlation	.011	1	.370
	Sig.(2-tailed)	.650		.000
	N	1615	1615	1615

doi: 10.18282/l-e.v10i1.2122

review-number	Pearson Correlation	-.169	.370	1
	Sig.(2-tailed)	.000	.000	
	N	1615	1615	1615

Table 3 Baby pacifier correlation

		star-rating	helpful-votes	review-number
star-rating	Pearson Correlation	1	-.070**	-.107**
	Sig.(2-tailed)		.000	.000
	N	18939	18939	18939
helpful-votes	Pearson Correlation	-.070**	1	.218**
	Sig.(2-tailed)	.000		.000
	N	18939	18939	18939
review-number	Pearson Correlation	-.107**	.218**	1
	Sig.(2-tailed)	.000	.000	
	N	18939	18939	18939

It can be seen from Tables 1, 2, and 3 that the correlation coefficients between the respective variables are very small, and the absolute values are all below 0.2. It can be considered that there is no correlation between the variables, and all can be included in the research model for analysis. In addition, it can be seen that the respective variables and control variables have significant correlations with the dependent variables (both of which are significant at the two-tailed level).

Therefore, it is analyzed that a hair dryer market and a baby pacifier market have a negative correlation with helpful-votes in favor of comments. There is a positive correlation between review-number and helpful-votes in favor of comments. There is a positive correlation between the star-rating and review-number and helpful-votes in a microwave oven market.

A preliminary test has been made through the correlation, and the specific impact relationship and its impact relationship under different commodities are discussed further through the subsequent regression analysis to obtain the support of each research hypothesis.In this paper, multiple linear regression analysis is used to verify the impact of the number of reviews and star reviews on the usefulness of reviews.

First, for all sample data, multiple linear regression analysis was performed between the respective variables and the usefulness of the reviews. Due to the non-normal distribution of star-rating, review- number, and helpful-votes, the review-number and helpful-votes are taking logarithmic, and review- number are standardized, and outliers are removed. In addition, in order to study the impact of star-rating on the usefulness of reviews, two variables, star-rating and star-square, are introduced. If the coefficients of star-term are negative and the coefficients of star-square term are positive, it indicates a "U" -shaped relationship. Otherwise it is an inverted "U" relationship; The final analysis results are shown in tables 4 and 5.

Table 4 Summary of hair dryer model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.260a	.067	0.67	12.0719	1.841

Table 5 Hair dryer ANVOA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	102782.887	2	51391.444	352.649	.000b
	Residual	1422759.079	9763	145.730		
	Total	1525541.967	9765			

Table 4 gives the overall fitting results of the model. It can be seen from the table that the adjusted judgment coefficient is 0.67, which has reached the level of similar research. It can be seen from Table 5that the observed value of the F statistic is 145.730, and the corresponding probability P value is approximately 0, reaching a significant level, indicating that it is valid and statistically significant, that is, the explanatory power of the independent variables in the model reaches a significant level.

Table 6 Hair dryer regression

Model

Unstandardized Coefficients B Std. Error

Standardized Coefficients

Beta t Sig.

Collinearity Statistics Tolerance VIF

1	(Constant)	-.905	.453		-2.000	.046		
	star-rating	-.026	.099	-.003	-.261	.794	.989	1.011
	review-number	.056	.002	.259	26.387	.000	.989	1.011

Table 7 Microwave oven regression

Model

Unstandardized Coefficients

B Std. Error

Standardized Coefficients

Beta t Sig.

Collinearity Statistics Tolerance VIF								
1	(Constant)	-5.860	1.627		-3.602	.000		
	star-rating	1.280	.395	.076	3.241	.001	.972	1.029
	review-number	.082	.005	.383	16.354	.000	.972	1.029

Table 8 Baby pacifier regression

Model	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
1	(Constant)	.814	.160		5.081	.000		
	star-rating	-.228	.034	-.047	-6.638	.000	.989	1.012
	review-number	.020	.001	.213	29.849	.000	.989	1.012

From table 6, table 7, and table 8, it can be seen that the coefficient of review-number is positive and significant at a significance level of 0.01, indicating that the review-number is significantly positively correlated with the star-rating. The star-rating coefficient in Table 6 is -0.026, the star-rating coefficient in Table 7 is 1.280, and the star-rating coefficient in Table 8 is -0.228, which is significant at the level of 0.001, indicating that the usefulness of star-rating is "U" Shape relationship.

3 Conclusion

Through the establishment of the above model and analysis of the data, we conclude that there is a negative correlation between the hair dryer market and the baby pacifier market and the number of comments and approvals, and the number of comments and approvals have a positive correlation; while the microwave oven market has star ratings and comments.There is a positive correlation with the number of comments and approvals, and it will become more and more significant over time.

References:

[1] Chatterjee P. Online Reviews: Do Consumers Use Them? Advances in Consumer Research[J],2011,28(5).

[2] Mudambi S M, Schuff D. What Makes A Helpful Online Review? A Study of Customer Reviews on Amazon.Com [J]. MIS Quarterly, 2010,34(1).

[3] Forman C, Ghose A, Wiesenfeld B. Examining. The Relationship Between Reviews And Sales: The Role of Reviewer Identity Disclosure in Electronic Markets [J]. Information System Research, 2008,19(3).

[4] Yin Guopeng. What kind of online reviews do consumers think are more useful? —— The effect of social factors [J]. Management World, 2012, (12).

[5] Wang Junkui. Research on the Usefulness of E-commerce Website Online Reviews[J]. Xi dian University,2014,3.

[6] Tan Yunzhi, Zhang Min, Liu Yiqun, Ma Shaoping. Collaborative Recommendation Framework Based on Ratings and Textual Reviews [J]. Pattern recognition and artificial intelligence, 2016-29(4).

[7] Li Huiying. Online reviews of consumer perceptions and business product sales affect research [J]. Harbin Institute of Technology,2013,7.